The Virtual Data Toolkit A Common Approach.

News Update – 5 Mar 2004

Update: May 6th 2005

Grid Physics Network

Particle Physics Data Grid

International Virtual Data Grid Laboratory

the Trillium Collaboration & Open Science Grid Consortium

PPDG Middleware Hardening

A focus of the Particle Physics Data Grid (PPDG) project is to integrate, extend and harden middleware for distributed computing with the end to end applications of several experiments in high energy and nuclear physics. PPDG, together with the NSF funded GriPhyN and iVDGL projects (the Trillium consortium), has adopted the use of the VDT packaging of grid middleware and the enhancements, debugging and bug fixing resulting from this broad deployment across many particle, nuclear and astrophysics experiments is benefiting a much wider science community.

What is the VDT?

The Virtual Data Toolkit (VDT) is an ensemble of grid middleware that can be easily installed and configured. VDT was originally created to serve as delivery channel for grid technologies developed and hardened by the NSF funded GriPhyN and iVDGL projects. These two projects are the primary sources of funding for VDT activities. However, the role of VDT has expanded over the last year and now supports the LHC Computing Grid Project (LCG) and the Particle Physics Data Grid (PPDG). U.S. CMS and U.S. ATLAS were early adopters of VDT and played an important role in enhancing the functionality and robustness of the toolkit components. Grid3 has provided a 35-site scale installation of a common version of the toolkit. All Experiments on PPDG use the VDT.

What Components are in VDT?

VDT contains three kinds of middleware:

- **Basic Grid services:** Includes Condor-G and the Globus Toolkit.
- **Virtual Data Tools:** Includes software tools developed by the Griphyn project to work with virtual data, particularly the Virtual Data System.
- **Utilities:** Includes a variety of tools: GSI-Enabled OpenSSH, software to update GSI certificate revocation lists, monitoring software like MonaLisa.

The PACMan utility is used to package the toolkit..

NMI Build and Test Facility

Most VDT components are built using machinery (software and computers) developed and deployed by the NSF National Middleware Initiative (NMI). The software packages use the Grid Packaging Toolkit (GPT) utility to build software bundles on the NMI build pool, which are then given to VDT.
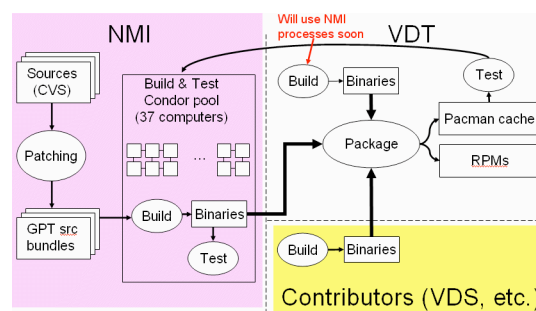
Middleware Benefits from PPDG

Extensions and robustness improvements in the Condor DAGMAN software, developed through the support of PPDG experiments, is now benefiting the biology community at the University of Wisconsin in their execution of BLAST. They have been able to increase the number of comparisons per run from the millions to over 4 billion. For the CNS/Cyana group successful computational runs have increased from several thousand to over 25000 CPU hours.

Accomplishments for GADU, the Genome Analysis and Database Update system, benefited directly from several key deliverables of PPDG contributions to the Grid services on which the GADU system relies. More than 10M genome sequences were processed by GADU on Grid3 resources at a throughput more than 5 times faster than the pre-Grid capabilities of this tool.

Storage Resource Management (SRM) middleware technology was applied to the Earth Science Grid (EDG), and adapted to work with a legacy system at NCAR. It is now deployed in several institutions for use by ESG projects, including ORNL, NERSC, NCAR, and LBNL.